Successful Adaptation of a Group EEG Generative Model to Individuals Depends on Participant Typicality

Xinyu Li [1], Marieke K. van Vugt [2], Natasha M. Maurits [1]

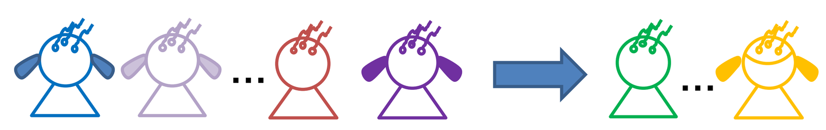
[1] University of Groningen, University Medical Center Groningen (UMCG), Department of Neurology, Groningen, The Netherlands.

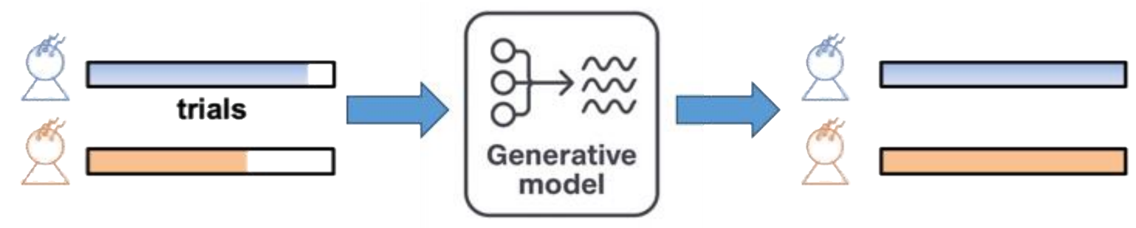
[2] University of Groningen, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, Groningen, The Netherlands.

Introduction

Machine learning can be used to identify patterns in EEG data. However, sometimes **there is insufficient data to do this.**

- **Case1:** not enough patients: **generate artificial EEG data** at the individual level
- **Case 2:** not enough trials per participant: **generate artificial EEG data** at the trial level





Problem: how to generate participant-specific EEG trials for different conditions with individual physiological plausibility?

Methods

1 Dataset and preprocessing

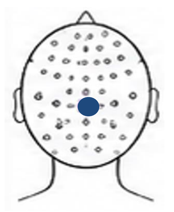
Public EEG dataset [1]

visual search task

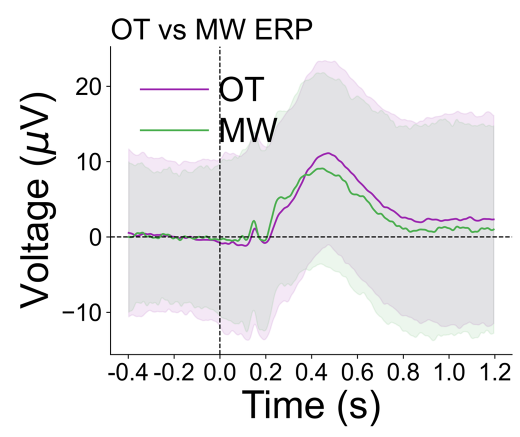
thought-probe labels: OT vs. MW

OT: 4753 trials (25 eligible participants)

MW: 1789 trials (12 eligible participants)



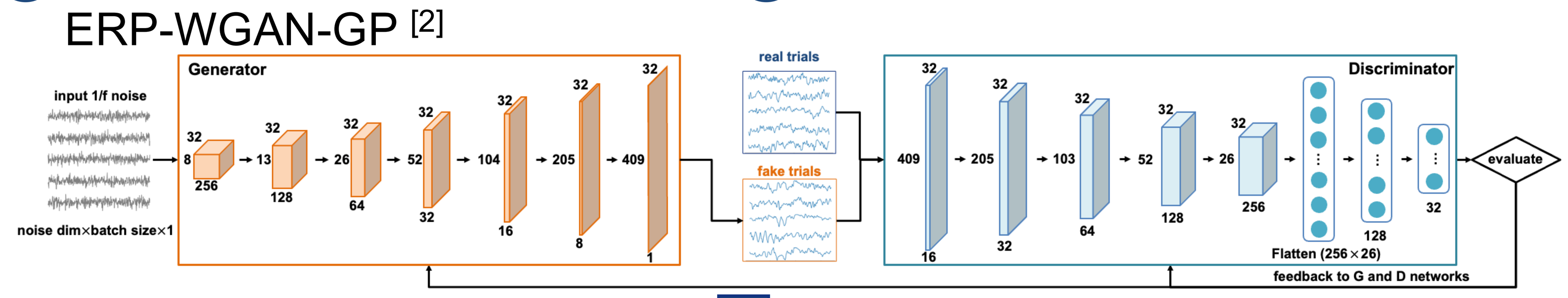
- Single channel (near Pz)
- 256 Hz sampling rate
- 1600 ms trials
- Re-reference (average)
- 0.5-40 Hz band-pass filter
- ICA artifact removal

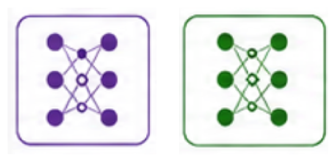


2 Group-pretrained models

ERP-WGAN-GP [2]

3 Participant-wise full fine-tuning





Full fine-tuning (all parameters updated)

4 Fake trial evaluation

(1) waveform inspection; (2) ERP and PSD plot evaluation;

(3) extracted EEG feature evaluation;

Real trials / Fake trials (the same number) → EEG feature extraction:
1~2. P3 (latency & amplitude)
3~5. Hjorth parameters (activity, mobility, and complexity)
6. Petrosian fractal dimension (PFD)
7~11. delta, theta, alpha, beta, and gamma power spectra
12~16. delta, theta, alpha, beta, and gamma entropy

→ Real features / Fake features (each trial) → KS test

→ PCA and t-SNE (visualization)

(4) quantitative similarity metrics (Real vs. Fake)

- $PCA\text{-}overlap_{RF}, PCA\text{-}BC_{RF}, PCA\text{-}JS_{RF}, SWD, MMD\text{-}RBF, WD$

Fine-tuning quality was based on $PCA\text{-}overlap_{RF}$:

Good: ≥ 0.60; Borderline: 0.50-0.60; Poor: < 0.50.

5 Participant-level predictor analysis

Predictors computed from **real data** only

Predictors (per participant)

- 16 feature deviations
- 2D PCA_{PG} centroid distance
- 16D centroid distance
- ERP deviation
- PSD deviation
- Mahalanobis distance
- Local density

Spearman correlation

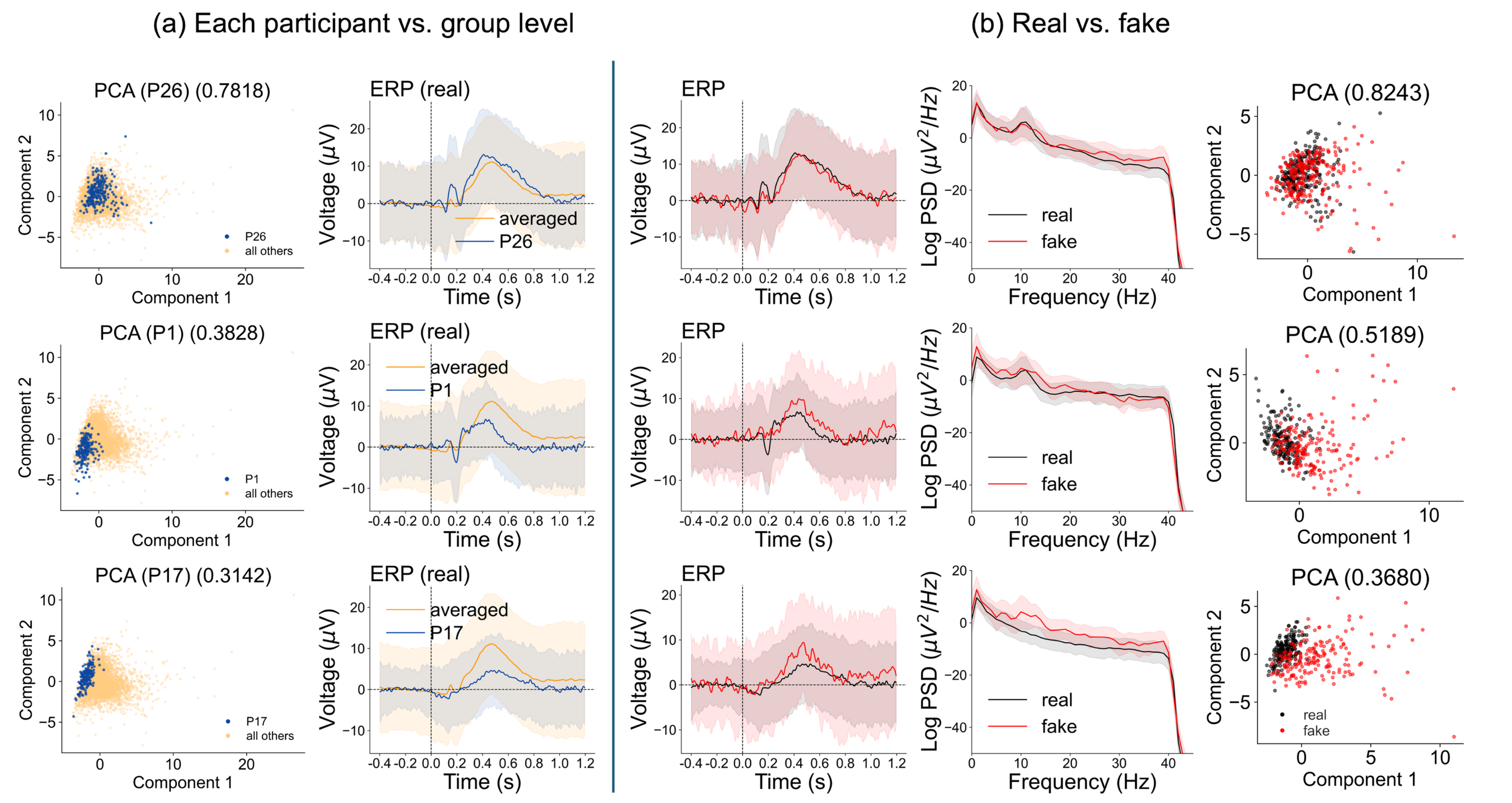
Outcome: $PCA\text{-}overlap_{RF}$

Metric Abbreviation

BC: Bhattacharyya coefficient
JS: Jensen-Shannon divergence
MMD-RBF: maximum mean discrepancy with radial-basis-function kernel
PCA: principal component analysis
PG: participant-group; **RF:** real-fake
SWD: sliced Wasserstein distance
WD: time-point-wise Wasserstein distance

Results

OT examples linking PG real-data deviation and RF quality



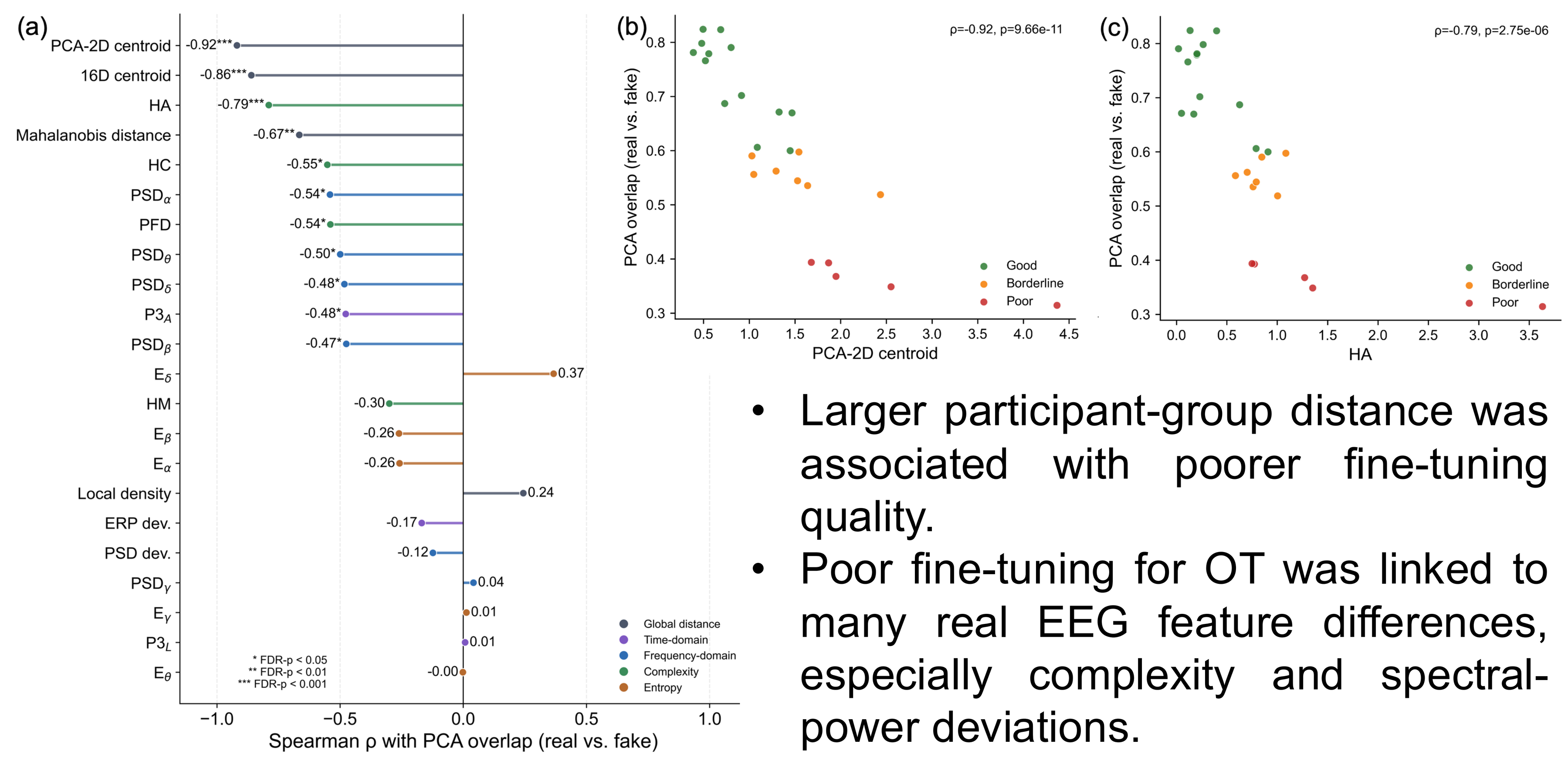
- Participants differ in how well the group-level model transfers.
- Participant 26 (P26), closer to the group-level feature distribution center, shows a good real-fake ERP and PS match.
- Participant 17 (P17), at the edge of the group-level feature distribution, shows a poor real-fake ERP and PS match.

Fake-trial quality metrics across participants for OT and MW fine-tuning

Task condition	n	PCA-overlap$_{RF}$ ↑	PCA-BC$_{RF}$ ↑	PCA-JS$_{RF}$ ↓	SWD↓	MMD-RBF↓	WD↓
OT	25	0.6090 ± 0.1575	0.8394 ± 0.1083	0.1402 ± 0.0911	3.0133 ± 0.9194	0.0225 ± 0.0126	2.5862 ± 0.9273
MW	12	$0.6887 \pm 0.0534^*$	$0.9102 \pm 0.0232^*$	$0.0831 \pm 0.0214^*$	2.4985 ± 0.6446	$0.0141 \pm 0.0050^*$	2.0966 ± 0.7617

Values are reported as mean ± standard deviation. arrows: direction of better quality, n: number of eligible participants. Welch's two-sample t-test was used for OT-MW fine-tuning comparison. Benjamini-Hochberg false discovery rate-corrected significance level: $*\ p < 0.05$.

Why fine-tuning quality varies (only OT)



- Larger participant-group distance was associated with poorer fine-tuning quality.
- Poor fine-tuning for OT was linked to many real EEG feature differences, especially complexity and spectral-power deviations.

Conclusion

- Group-pretrained EEG generative models (ERP-WGAN-GP) can be fine-tuned to individual participants.
- Fine-tuning is not always successful: MW showed stable real-fake agreement, whereas OT showed larger participant-level variability.
- For the OT condition, larger participant-to-group real EEG deviation is associated with poorer fine-tuning quality.
- Participant-to-group deviation may help identify when individualized EEG generation is feasible.

LinkedIn

rijksuniversiteit groningen | umcg

Email: x.li01@umcg.nl

Abbreviation:
EEG: electroencephalogram; **ERP**: event-related potential
G: Generator; **D**: Discriminator; **WD**: Wasserstein distance
ICA: independent component analysis;
KS test: Kolmogorov-Smirnov test
MW: mind-wandering; **OT**: on-task; **PSD**: power spectral density
t-SNE: t-distributed stochastic neighbor embedding
WGAN-GP: Wasserstein generative adversarial networks with gradient penalty

References:

[1] Jin, Christina Yi, Jelmer P. Borst, and Marieke K. Van Vugt. "Predicting task-general mind-wandering with EEG." Cognitive, Affective, & Behavioral Neuroscience 19 (2019): 1059-1073.

[2] Li, X., van Vugt, M. K., & Maurits, N. M. "Evaluating the impact of input noise and ERP-based penalties on the physiological plausibility of EEG generation using WGAN-GP." Computers in biology and medicine 199 (2026): 111296.

Neurologie GRONINGEN